Detecting Fake News on Social Media: A Machine Learning Approach

Rajat Nagar
rajatnagar@vt.edu
Virginia Tech
Falls Church, USA

Sanjna Kumari
sanjnak@vt.edu
Virginia Tech
Falls Church, USA

Shivani Patil
shivanidinkar@vt.edu
Virginia Tech
Falls Church, USA

ABSTRACT

The rapid spread of fake news on social media platforms has significant consequences for society, influencing public opinion, election outcomes, and economic markets. This project develops machine learning models to classify news statements as real or fake using textual content. Here we will present the linguistic characteristics of news statements that have gone viral on social media by providing design tools to identify and contain the propagation of misinformation on the web. We used the PolitiFact dataset, simplifying its labels into binary classes, while we explore various machine learning algorithms along with feature extraction techniques to determine the most effective approach in detecting fake news.

KEYWORDS

Fake News Detection, Social Media Analysis, Machine Learning, Text Classification, Natural Language Processing

1 INTRODUCTION

The digital age has revolutionized how information is disseminated and consumed, with social media platforms playing a pivotal role in the rapid spread of news. However, this ease of information sharing has also led to the proliferation of fake news—false or misleading information presented as news—which poses real consequences for society. Fake news can influence how people think, sway election outcomes, and impact economic markets. For instance, fake stories about Hillary Clinton circulated extensively during the 2016 U.S. presidential election, potentially influencing voter opinions. Another example is a fake news report about an explosion injuring former President Barack Obama, which led to a loss of $130 billion in stock value.

Given the speed and low cost at which information spreads online, there is an urgent need for processes that can identify and alert users against fake news. This project aims to develop machine learning models to classify news statements as real or fake using textual content, thereby assisting in identifying and containing the spread of misinformation on the web.

2 RELATED WORK

The detection of fake news has become a critical area of research in natural language processing and machine learning. Previous studies have employed various techniques ranging from linguistic cue analysis to deep learning models. Wang [1] introduced the LIAR dataset, a benchmark dataset for fake news detection, which includes detailed labels and metadata. Other research has explored the use of ensemble methods and neural networks to improve classification accuracy [2, 3].

While significant progress has been made, challenges remain due to the complexity and subtlety of language used in fake news. Our work builds upon these foundations by utilizing the PolitiFact dataset and experimenting with multiple algorithms and feature representations to enhance detection performance.

3 DATASET

3.1 Source

The dataset used in this project is derived from PolitiFact, as described in the paper "Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection" by Wang (2017) [1]. PolitiFact is a Pulitzer Prize-winning fact-checking organization that evaluates the truthfulness of statements made by politicians and public figures.

3.2 Key Attributes

The dataset includes the following key attributes:

- **ID**: Unique identifier for each statement.
- **Label**: Original labels include 'true,' 'mostly-true,' 'half-true,' 'false,' 'barely-true,' and 'pants-fire.'
- **Statement**: The news statement to be classified.

3.3 Simplification

For the purpose of this project, the six original labels have been simplified into two classes to facilitate binary classification:

- **True Class**: Contains 'true,' 'mostly-true,' and 'half-true' statements.
- **False Class**: Contains 'false,' 'barely-true,' and 'pants-fire' statements.

3.4 Examples from the Dataset

3.4.1 *False News Statements.*

- "Says the Annies List political group supports third-trimester abortions on demand."
- "Health care reform legislation is likely to mandate free sex change surgeries."
- "Jim Dunnam has not lived in the district he represents for years now."

3.4.2 *True News Statements.*

- "When did the decline of coal start? It started when natural gas took off that started to begin in (President George W.) Bush's administration."
- "The economic turnaround started at the end of my term."

- "I'm the only person on this stage who has worked actively just last year passing, along with Russ Feingold, some of the toughest ethics reform since Watergate."

These examples illustrate the complexity and variety of statements and highlight the challenges in correctly classifying them using text alone.

4 METHODOLOGY

We conduct experiments using various machine learning algorithms and feature extraction techniques to identify the most effective approach for fake news detection.

4.1 Algorithms

- **Naive Bayes (MultinomialNB)**: Suitable for text classification with discrete features, leveraging the assumption of feature independence.
- **Logistic Regression**: A linear model effective for binary classification tasks.
- **Support Vector Machines (SVM)**: Particularly the linear SVM, which performs well in high-dimensional spaces.
- **Random Forest**: An ensemble method that builds multiple decision trees and merges them to improve prediction accuracy and robustness.

4.2 Feature Representations

- **Binary Weighting**: Represents the presence or absence of words in the statements.
- **Count Vectorization**: Statements are represented through the frequency of each word.
- **Term Frequency-Inverse Document Frequency (TF-IDF)**: Weighs words based on their importance, emphasizing those that carry more information by considering their frequency across all documents.

4.3 Preprocessing Steps

- **Text Preprocessing**: Conversion of all text to lowercase, removal of punctuation, special characters, and stopwords to reduce noise.
- **Tokenization**: Breaking down statements into constituent words or tokens for analysis.

5 EXPERIMENTAL SETUP

5.1 Data Preparation

The dataset is split into training and testing sets, ensuring a balanced distribution of true and false statements. Preprocessing steps are applied to clean the data and prepare it for feature extraction.

5.2 Training and Evaluation

Each algorithm is trained using different feature representations. We perform hyperparameter tuning using cross-validation to optimize model performance.

5.3 Metrics

The primary evaluation metric is the **F1-score**, which balances precision and recall:

- **Precision**: The accuracy of positive predictions (i.e., the proportion of statements classified as fake that are actually fake).
- **Recall**: The ability of the model to find all relevant instances (i.e., the proportion of actual fake statements that were correctly identified).

Using the F1-score is crucial because both false positives (misclassifying real news as fake) and false negatives (failing to identify fake news) have significant consequences. We also utilize the confusion matrix to analyze the types of errors made and to fine-tune our models accordingly.

6 EXPECTED OUTCOMES

We anticipate that ensemble methods like Random Forest may perform better due to their ability to handle complex patterns in the data. Feature representations like TF-IDF are expected to improve model performance by emphasizing informative words. Through comparative analysis, we aim to identify the most effective combination of algorithms and feature extraction techniques for fake news detection.

7 ETHICAL CONSIDERATIONS

Detecting fake news involves ethical considerations, particularly concerning censorship and freedom of speech. Our model aims to assist in identifying misinformation without infringing on individuals' rights to express opinions. We acknowledge the potential biases in data and strive to minimize them through careful preprocessing and validation. The goal is to support fact-checking efforts and promote informed decision-making in society.

8 CONCLUSION

This project explores machine learning approaches to detect fake news on social media by classifying news statements as real or fake. By experimenting with multiple algorithms and feature representations, we seek to enhance the accuracy of fake news detection models. The outcomes have the potential to contribute significantly to efforts aimed at mitigating the spread of misinformation online.

9 REFERENCES

[1] Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. *arXiv preprint arXiv:1705.00648.*

[2] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.

[3] Vaidya, A., Mai, T., & Ning, P. (2016). Empirical Analysis of Machine Learning Approaches for Fake News Detection. *Proceedings of the 2016 International Conference on Collaboration Technologies and Systems (CTS)*, 111-118.